

PREDICTION OF RESIDENTIAL MORTGAGE CONTRACT RATES

by

FENGLING YE

(Under the direction of Jaxk Reeves)

ABSTRACT

In this thesis, we have examined the predictability of the effective yield of residential mortgage loan rates as a function of economic variables which would be available to a loan officer preparing to set a loan rate. Of the possible explanatory regressor variables, the 10-year interest rate is most highly correlated with effective yield. Since this is a case where the sample sizes (number of loans made per month) vary widely, the major new statistical idea introduced in this thesis is that of the weighted autoregressive first order model [WAR(1)]. The procedure used in this thesis to obtain the [WAR(1)] estimates is based on iteratively reweighted least squares methods.

INDEX WORDS: Weighted Autoregression Model; Autoregressive Error Model;
 Effective Yield.

PREDICTION OF RESIDENTIAL MORTGAGE CONTRACT RATES

by

FENGLING YE

B.E.C., The University of International Business and Economics, China 1986

M.A., The University of International Business and Economics, China 1989

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2002

© 2002

Fengling Ye

All Rights Reserved

PREDICTION OF RESIDENTIAL MORTGAGE CONTRACT RATES

by

FENGLING YE

Approved:

Major Professor: Jaxk Reeves

Committee: Nancy Lyons
William McCormick

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2002

ACKNOWLEDGMENTS

I am truly grateful to Dr. Jaxk Reeves, my major professor, whose detailed instructions and selfless sacrifices of personal time have inspired me profoundly. This thesis would have never been completed without his knowledge and encouragement.

I am very thankful to Dr. Nancy Lyons and Dr. William McCormick, my committee members, for their timely assistance and insightful comments.

I am also indebted to my employer and my dear colleagues, who supported me and my work with their love and accommodations.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	vii
CHAPTER	
1 INTRODUCTION	1
2 DISCUSSION OF THE MEAN AND SD OF EFFECTIVE YIELD	4
3 STATISTICAL ANALYSIS	12
3.1 REGRESSION APPROACH	12
3.2 PURE AR(1) APPROACH	14
3.3 AUTOREGRESSIVE ERROR MODEL	16
4 WEIGHTED AUTOREGRESSIVE MODEL	25
5 CONCLUSION	32
BIBLIOGRAPHY	36

LIST OF FIGURES

2.1	Weighted Mean of Effective Yield for Four Point Strata	7
2.2	Weighted Mean of Effective Yield by Year	8
2.3	Weighted Mean of Effective Yield by Month	9
2.4	Effective Yield of the Four Point Strata	10
2.5	Plot of Monthly Effect Adjustments	11
3.1	Residual*CDATE plot of WLS Model, Strata (-1)	20
3.2	Residual*CDATE plot of WLS Model, Strata (0)	20
3.3	Residual*CDATE plot of WLS Model, Strata (1)	21
3.4	Residual*CDATE plot of WLS Model, Strata (2)	21
3.5	Residual*CDATE plot of AR(1) Model, Strata (-1)	22
3.6	Residual*CDATE plot of AR(1) Model, Strata (0)	23
3.7	Residual*CDATE plot of AR(1) Model, Strata (1)	23
3.8	Residual*CDATE plot of AR(1) Model, Strata (2)	24

LIST OF TABLES

2.1	Mean & SD Values of Effective Yield By Strata	7
2.2	Weighted Mean of Effective Yield by Year	8
2.3	Weighted Mean of Effective Yield by Month	9
2.4	Mean Effective Yield By Point Strata	10
2.5	Monthly Adjustment for Effective Yield	11
3.1	Correlation Coefficients by Point Strata	18
3.2	P-Values of Full and Reduced WLS Models (All Strata Combined)	19
3.3	Weighted Least Squares Regression Model Parameter Estimates	19
3.4	Parameter Estimates of Pure AR(1) Model (Separate Strata)	19
3.5	Fit of AR(1) Difference Model	22
3.6	Parameter Estimates of AR(1) Error Model	22
4.1	Estimates of σ_y , σ_a and ϕ for WAR(1) Model	30
4.2	Parameter Estimates for WAR(1) Model, $\sigma_a = .003$	31
4.3	Estimates of β_0 , β_1 and RMSE for WAR(1)E Model	31
5.1	Comparison of RMSE for Five Different Models	34

CHAPTER 1

INTRODUCTION

Mortgage loan rates are affected by many national factors, especially economic factors. Current economic conditions influence the short-term loan interest rate the most. How should banks decide on the 30-year fixed-rate residential mortgage loan rates? A bank in California has conducted a study on this, in which they recorded information from all residential 30-year fixed-rate mortgage loans during the years 1991 to 2000. This information was recorded monthly and separately for each “point strata”, as explained below. The data was given under a confidential agreement to a faculty member in UGA’s Terry College of Business. Neither the bank, the faculty member, nor the graduate student who supplied us with this data were willing to comment further on the data for the fear that a competitor might somehow use this data to gain competitive advantage. Considering what we will discover later in this thesis about the data, that seems quite unlikely.

The data set consists of 480 observations (4 strata*120 months), with the following recorded for each observation:

Y: Effective Yield = Contract Rate + Points; averaged over all loans within the strata;

YR10: 10-year T-Bill Rate averaged over days in month;

MR3: 3-month T-Bill Rate averaged over days in month;

VOL: volatility; the SD of daily 10-year rate over last 7 months, including current month;

PT: percentage financial charge by the bank: -1%, 0%, 1%, 2%;

N: number of loans from which the effective yield is calculated for the month;

MON: month (January, February, December).

YR: year (from 1991 to 2000).

The effective yield (Y) of a loan is the rate quoted to the customer after adjusting for “PT”(Point). It is contract rate plus PT (points) of the loan. This is the primary response variable of interest. We would like to predict what the effective yield will be, based on the information which would be available to a loan manager at the time a rate is set. The economic variables YR10 (10-year interest rate), MR3 (3-month interest rate), and VOL (7-month volatility) are obtained from US Government data. Since many economists use the difference between the mid-term interest rate and the short-term interest rate as an indicator of market consumption, we have also considered a new variable “TERM” in the following analysis ($TERM = YR10 - MR3$).

The “PT” (point) is a type of financial charge which is usually positive, though it could be zero or even negative, to entice a customer and to compete with other banks (Clauret, 1999). This is the stratification variable used to separate the loans into four types. The PT in this data set are: -1%, 0%, 1% and 2%. For most months in the 10-year series there are a reasonable number of observations in each strata. The average sample size per month for the four strata are 124, 228, 73 and 43, respectively. However, in some months, there are very few loans in certain point strata. Most loans’ point values do fall exactly on one of these integer percent values, but for the few which did not, the loan was included in the nearest integer strata.

The variables MON (month) and YR (year) are fairly straight-forward, and can be combined to order the series. To do this, we created the variable “CDATE”, which counts

the months as -60, -59, . . . , 0 . . . , +59, +60. This is useful in determining whether a trend is present.

There are four time series in the data, each of 10 years' (120 months) duration. Each observation represents a kind of home mortgage loan based on the PT associated with the loan. For the 120 months in each of the four time series, information on all the eight quantities listed above are recorded. The Y (effective yield) variable is a rate averaged over all loans of the type made by the bank in a particular month. For many of the months in these series, the number of loans per strata is near the grand mean of 118 loans, though it could be as high as $n=1257$, or as low as $n=0$.

The primary point of interest here is to see if the effective yield of residential mortgage loan rates can be predicted and, if so, to see how it is related to the above-listed variables. One question of particular interest to the economists who obtained this data set is whether residential mortgage rate is related to the mid-term interest rate (YR10) and/or to short-term interest rate (MR3). Although this would appear to be a routine statistical analysis, there are enough complicating factors present to warrant a more thorough investigation. Among these are: unequal sample sizes, stratification issues, correlated measurements, and missing data. This thesis begins by assuming the simplest possible model and proceeds, by gradually accounting for various complications, until a realistic model is developed in Chapter 4. The procedure described herein would have applications in other commonly occurring situations, as discussed in Chapter 5.

CHAPTER 2

DISCUSSION OF THE MEAN AND SD OF EFFECTIVE YIELD

Even describing simple statistics for effective yield (Y) for each stratum is not necessarily straight forward. There are n=120 observations for each strata (one per month), but the number of loans observed per month (N) is highly variable, so a weighted average would be more appropriate to use. The difference between using unweighted and weighted observations for the Y variable in each stratum is illustrated in Table 2.1. One can easily note, whether using unweighted or weighted means, that as the PT value increases, the effective yield increases (Figure 2.1). This behavior is expected, although the approximately linear relationship observed is not really expected. From Figure 2.1, we observe that the slope of the line is approximately .0018 per point. One might have expected a slope of .01, since effective yield is defined to be contract rate plus points. However, that would be true only if the contract rate were the same for each strata. Clearly, the rates are set so that effective yields are almost equal; those with higher points get lower contract rates and vice versa. However, as Table 2.1 and Figure 2.1 show, the effective yield for fixed time is not quite constant; those in higher risk groups are still charged a little more. For most analyses, we will examine each strata separately, combining them in cases where pooling or linear shifts appear appropriate.

A very crucial point in this entire thesis concerns what the “SD” of Y is. The sample estimate of the SD, if we had every Y_{ij} (effective yield of loan j in month i), would, of course, be:

$$S = \sqrt{\frac{\sum_{i=1}^{120} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2}{N - 1}}$$

where $\hat{\mu}$ is the weighted mean,

$$\hat{\mu} = \frac{\sum_{i=1}^{120} \sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^{120} \sum_{j=1}^{n_i} 1} = \frac{\sum_{i=1}^{120} \sum_{j=1}^{n_i} Y_{ij}}{N}.$$

With some algebraic manipulation, one can rewrite the expression within the radical, the sample variance, as

$$S^2 = \frac{\sum_{i=1}^{120} n_i \hat{\mu}_i^2 - N \mu^2 + \sum_{i=1}^{120} (n_i - 1) S_i^2}{N - 1} = \sigma_B^2 + \sigma_a^2,$$

where $\hat{\mu}_i$ and S_i are the sample mean and sample SD of Y within each of the 120 months. Unfortunately for us, the sample SD's (S_i) for each month were not recorded in the original data set, and obtaining them at this point is nearly impossible because of the sensitivity of the data. The “Unweighted SD” in Table 2.1 assumes that each $n_i = 1$, so $N = 120$, which is clearly wrong. The “Weighted SD” in Table 2.1 uses the correct n_i values, so the “between-month” SD (σ_B) is estimated correctly, but the within group SD's (S_i), were set = 0, yielding a “within-month” SD (σ_a) of zero, which is also not correct. The “Weighted Adjusted SD” is calculated using the estimate that $S_i = \sigma_a = 0.003$ for all months and all strata. This estimate was obtained from an examination of a small subset of the PT=0 raw data. It is not particularly reliable, but is the only estimate which we have at this time. Later in the thesis, we shall examine the effect of this assumption upon our results.

One might also wonder whether there is any “year effect” or “month effect” in these data. To examine this crudely, we pooled the data over all four strata and calculated the weighted means for each year and for each month. The values by year are shown in Table 2.2 and Figure 2.2, while those by month are shown in Table 2.3 and Figure 2.3, respectively. There does not appear to be any yearly trend, although the first year (1991) is much higher than average and two of the latter years (1998 and 1999) are much lower than average, so that a naive application of linear regression would indeed suggest a negative trend. Separate examination by strata shows 1991 to be particularly high for strata -1 and 0, while 1998 and 1999 are particularly low for strata 1 and 2.

The monthly data are somewhat more interesting. The shape in Figure 2.3 suggests a cyclical trend with rates lowest in February, rising to a peak in June, and then slowly declining. Such cyclic patterns are often seen in weather data (river flows peak in April, for example, when the snow melts; while they are lowest in the dog days of August and September [Wahl et al., 1995]), but were not really expected here.

To see if the effect was real, we examined plots similar to Figure 3 separately for each of the four strata and noted that all four had similar shapes (see Table 2.4 and Figure 2.4). Figure 2.4 clearly shows the cyclic pattern of effective yield due to month of each of the four strata. For the $PT=-1$ group (called '3' in Figure 2.4), the peak is in June and lowest in January. For the $PT=0$ group, the peaks are in May and September and lowest in February. For the $PT=1$ group, the peak is in June and lowest in December. For the $PT=2$ group, the peak is in May and lowest in January. The peaks are near the middle of the year and the minima are all in the first or last months of the year. The plot and the mean information conclude that there is a cyclic pattern due to month effect. Apparently, those who set loan rates tend to inflate them during late Spring and early Summer and to lower them in other times of the year.

Using the LSMEANS option in SAS, controlling for year and strata effects, we estimated the effects on Y due to each month to be as shown in Table 2.5 and Figure 2.5. For this plot, we have arbitrarily set the "zero" value as September, since it is a typical median month. The monthly pattern is similar to, but smoother than, that shown in Figure 2.3. In the more sophisticated time-series analyses included later in this thesis, we de-trended our data by subtracting the monthly coefficients shown in Table 2.5 from the Y values given in the raw data.

Table 2.1: Mean & SD Values of Effective Yield By Strata

Point Value	Unweighted		Weighted			N(# of Loans)
	Mean	SD	Mean	SD	Adj. SD	
-1%	0.07741	0.00422	0.07464	0.00504	0.00586	14,624
0%	0.07944	0.00435	0.07622	0.00497	0.00580	27,359
1%	0.08069	0.00736	0.07850	0.00657	0.00721	8,802
2%	0.08177	0.00842	0.07989	0.00813	0.00865	4,977
Total	0.07982	0.00653	0.07649	0.00537	0.00615	55,762

Figure 2.1: Weighted Mean of Effective Yield for Four Point Strata

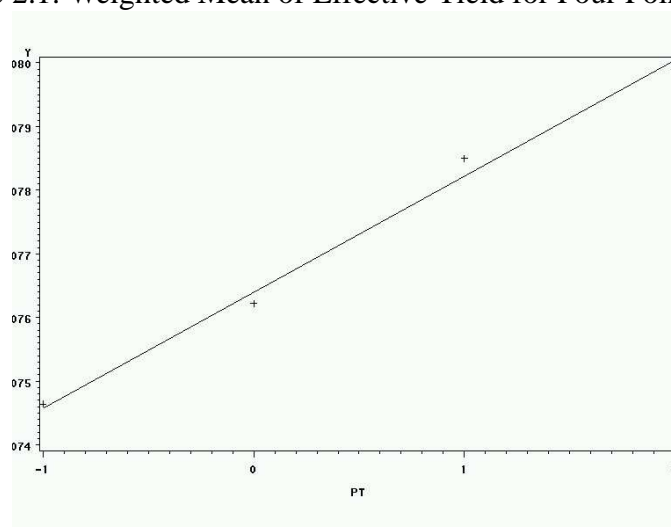


Table 2.2: Weighted Mean of Effective Yield by Year

YR	Mean of Y
1991	0.090657
1992	0.082749
1993	0.077927
1994	0.078096
1995	0.079601
1996	0.079316
1997	0.077837
1998	0.071563
1999	0.070231
2000	0.080779

Figure 2.2: Weighted Mean of Effective Yield by Year

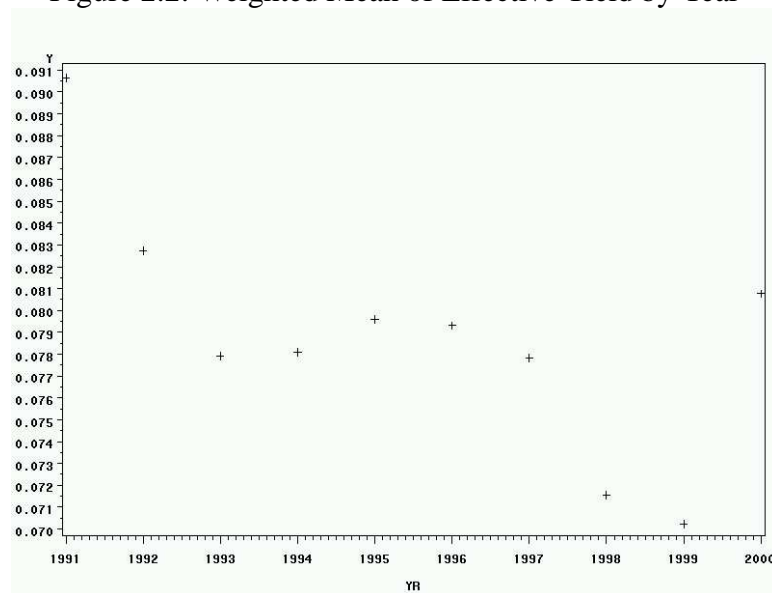


Table 2.3: Weighted Mean of Effective Yield by Month

MON	Mean of Y
1	0.075046
2	0.074695
3	0.075727
4	0.077330
5	0.077662
6	0.077801
7	0.077360
8	0.076740
9	0.076904
10	0.076457
11	0.076137
12	0.075732

Figure 2.3: Weighted Mean of Effective Yield by Month

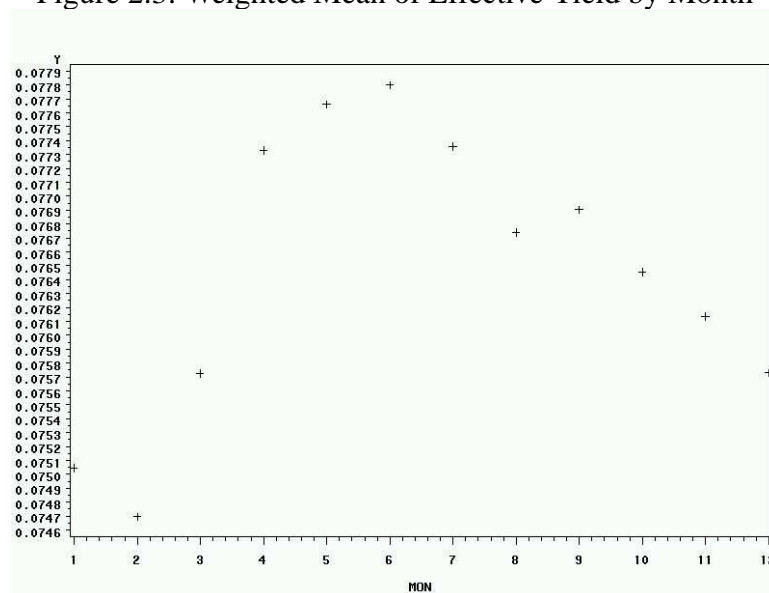


Table 2.4: Mean Effective Yield By Point Strata

POINT STRATA				
MON	(-1)	(0)	(1)	(2)
1	0.072425	0.075409	0.077461	0.076579
2	0.072489	0.074471	0.076926	0.077737
3	0.074328	0.075097	0.077781	0.078501
4	0.074974	0.077023	0.079160	0.082522
5	0.074768	0.077084	0.080567	0.084144
6	0.076358	0.076771	0.081893	0.082019
7	0.075951	0.076539	0.080932	0.081957
8	0.075595	0.075886	0.079580	0.080640
9	0.074987	0.077475	0.077694	0.078536
10	0.074908	0.076637	0.077483	0.078885
11	0.074371	0.076504	0.077425	0.077751
12	0.073835	0.076034	0.076777	0.078040

Figure 2.4: Effective Yield of the Four Point Strata

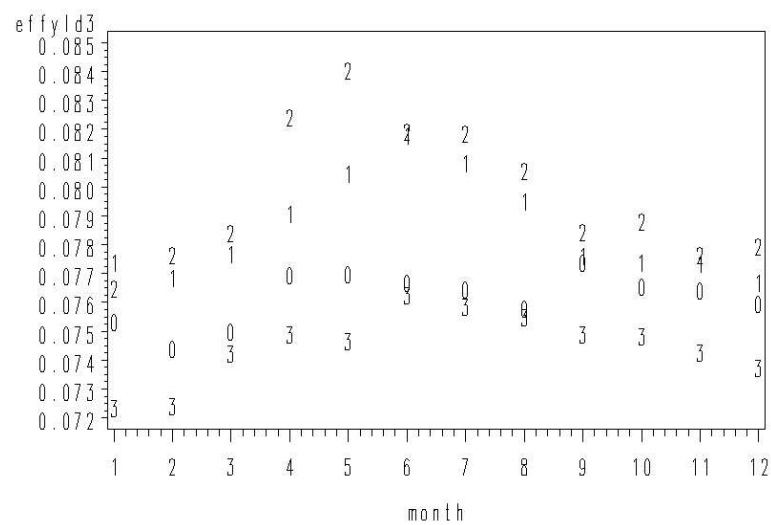
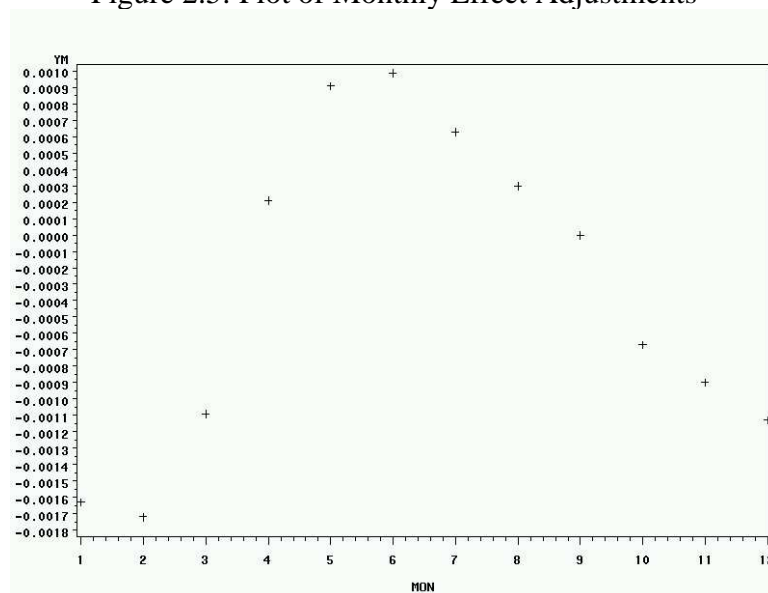


Table 2.5: Monthly Adjustment for Effective Yield

MON	Adjust.
JAN:	-0.00163
FEB:	-0.00172
MAR:	-0.00109
APR:	+0.00021
MAY:	+0.00091
JUN:	+0.00099
JUL:	+0.00063
AUG:	+0.00030
SEPT:	-0.00000
OCT:	-0.00067
NOV:	-0.00090
DEC:	-0.00113

Figure 2.5: Plot of Monthly Effect Adjustments



CHAPTER 3
STATISTICAL ANALYSIS

3.1 REGRESSION APPROACH

Although it was mentioned in Chapter 2 that it would be naive to apply linear regression to this data set, a regression study is carried out as a preliminary step. This is mainly for the purpose of deciding what factors are most important in prediction of effective yield and to see which factors should be included in the model. As a first step to answer this question, “PROC CORR” of SAS (SAS Institute, 1995) is performed to obtain the correlation coefficients of Y with key variables for each point strata as shown in Table 3.1.

In all the four strata, Y (effective yield) is most correlated with YR10, the 10-year interest rate. It is least correlated with MR3, the 3-month interest rate. Both interest rates, and their difference, TERM, are positively correlated with the response variable. Volatility is negatively correlated with the response variable. CDATE is negatively associated with Y, but as noted earlier, the trend is not really linear. Each of these variables separately appears to be a significant predictor of Y, but, since they are highly correlated with one another, it is not assured that they will be jointly significant.

Since $TERM = YR10 - MR3$, there is no point in attempting to include it in a model with YR10 and MR3. Thus, we might start with a full model of the form:

$$Y = \beta_0 + \beta_1 * (YR10) + \beta_2 * (MR3) + \beta_3 * (VOL) + \beta_4 * (CDATE) + error,$$

and see how much it can be reduced. The top half of Table 3.2 gives the output from “PROC REG” of SAS for the model above. This model is for all strata combined and

weights each observation by N , the number of loans over which Y is calculated. As expected, this model is over-parameterized. Employing various model selection strategies (stepwise selection and SBC), the following 2-variable model appears best by either criteria:

$$Y = \beta_0 + \beta_1 * (YR10) + \beta_2 * (VOL).$$

The parameter estimates for this reduced model are shown in the lower half of Table 3.2.

If one examines each of the strata separately, somewhat different best models emerge, as shown in Table 3.3. For all strata, YR10 is a very significant predictor, with variables such as MR3 and VOL sometimes adding significant predictability. Since this is a weighted regression, the “RMSE” calculated by SAS would represent the typical error observed with $n=1$ observations. The “typical RMSE” displayed in Table 3.3 is the RMSE which would be expected for a month where there are an average number ($n=N/120$) of loans observed.

Although, as mentioned previously, the errors are not independent and the assumptions for linear regression are violated here, one can still see from the above models that effective yield (Y) is very much associated with YR10 with no exception among the combined strata model and the separate strata models. YR10 seems to be the most important variable to predict Y . We can also conclude that each strata behaves differently; therefore we need a separate strata model for Y prediction. In addition, it should be noted that using $YR10_t$ or $MR3_t$ to predict Y_t , while feasible after the fact, would not be very useful practically. A fairer measure of the predictive ability of either YR10 or MR3 to predict Y would be to regress Y_t on $YR10_{t-1}$ and $MR3_{t-1}$. This is what was done in Table 3.3. While this decreased the explained variation (R-squared) in each strata slightly (as compared to regressing on $YR10_t$ or $MR3_t$), the resultant models are more realistic with respect to expected error, if utilized.

Figures 3.1-3.4 are residual plots for each strata, in each case plotting the residuals vs time (CDATE). There is some evidence that the early (1991-1992) residuals are more dispersed than those in other years, but the most noticeable defect of each of the four residual plots is that the errors are not at all independent. One sees extreme evidence of serial correlation, with positive residuals frequently followed by other positive residuals and vice-versa. This is a classical indicator that the temporal nature of recorded data has not been correctly accounted for. The next two sections of this chapter concern procedures to remedy this shortcoming.

3.2 PURE AR(1) APPROACH

Since both the response variable and the explanatory variables are from a time series, it would seem reasonable that one should use an autoregressive (AR) model to describe the relationship between successive observations. The evidence for doing so is especially strong after one views the RESIDUAL*CDATE plots in the previous section and notes how dependent successive errors are. It is almost obvious that one should use successive past records of Y to predict a future value of Y. The simplest model of this type is an AR(1) model (Shumway, 2000):

$$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + \epsilon_t,$$

where μ is the mean of the autoregressive process, and ϵ_t are assumed to be independent identically distributed normal mean = 0, SD = σ_ϵ (constant) random variables. For this process to be valid, the process should be stationary, so that estimating μ by the mean of Y will be reasonable.

In fact, as noted in Chapter 2, there appears to be a cyclic monthly effect and the mean levels appear to vary by strata, so a more accurate model to use would be:

$$(Y_t^J - \mu_t^J) = \phi(Y_{t-1}^J - \mu_{t-1}^J) + \epsilon_t^J,$$

where $\mu_t^J = \mu^J + \alpha_\nu$, $t = 12K + \nu$, $J = -1, 0, 1$ or 2 are the strata indices and $K = 0, \dots, 9$ are year indices. That is, there is a separate grand mean (μ^J) for each type of the four strata (with means as given by the “unweighted mean” column in Table 2.1), and there are twelve monthly effects, $\alpha_1, \alpha_2, \dots, \alpha_{12}$, independent of strata, with these twelve values estimated as shown in Table 2.5. Fitting these simple models to the four strata yield the parameter estimates shown in Table 3.4.

The residual plots (Figures 3.5-3.8) of these models look significantly better than those of the regression models in Section 3.1, indicating that AR(1) is probably sufficient. Partial correlation graphs (not shown) confirm that PACF values beyond lag 1 fall within the two standard errors bound, suggesting the possible suitability of an AR(1) model (as opposed to higher-order autoregressive models) for the mean corrected data set. Also, the RMSE's are greatly improved, especially in strata -1 and 0.

However, a number of problems remain. Among these are:

- a) All of the ϕ estimates are quite near +1. This is a cause for concern in AR(1) modelling, since $|\phi| > 1$ indicates an unstationary AR(1) process. If $|\phi|$ is near 1, many statistical analysts would consider instead the first difference series $Z_t = Y_t - Y_{t-1}$, trying to model Z_t as white noise. A white noise model for the difference series appears to be correct, with error SD's being estimated as shown in Table 3.5.

For strata -1 and 0, the RMSE's of the difference model are about the same as those given by AR(1) model (see Table 3.4). However, for strata 1 and 2, the difference model yields RMSE's which are about 1.5 times larger than those given by AR(1) model. This is somewhat surprising, since the autoregressive parameter estimates for strata 1 and 2 are very close to one (0.9833 and 0.9764), hence should correspond closely to the difference model.

- b) Although Y_t appears to be well-explained by Y_{t-1} , it may still be the case that $YR10_t$ or $YR10_{t-1}$ have significant explanatory power. This matter is examined in Section 3.3.
- c) The most important objection of all, dwarfing those in (a) and (b), is that these models treat the series as 120 equally valid observations. This is why the “unweighted means” are used to estimate the strata means. However, as noted in Chapter 2, the observed Y values are averages over N loans, and the values of N vary widely. There is no standard SAS routine to perform “weighted AR(1) analysis”. Remedying this problem is the focus of Chapter 4.

3.3 AUTOREGRESSIVE ERROR MODEL

As was noted in the regression analysis of Section 3.1, YR10 is highly correlated with Y and explains Y the best among the potential predictor variables. In the pure AR(1) model of Section 3.2, we have used only the information on the Y series, without including any other variables that might really be related to Y and potentially important in the prediction. Thus, another kind of model should be explored. This model is a combined regression and autoregression model, sometimes referred to as the Autoregressive Error Model.

The general idea is that the autoregressive error model will explain the data set better and give more precise predictions than either of the other two types. The WLS regression model, as noted in Section 3.1, does not account for correlated errors. The pure AR(1) model of Section 3.2 accounts for correlated errors, but not for other covariates. So, the Autoregressive Error Model will make efficient use of all the information provided in the data set. The AUTOREG procedure of SAS (SAS Institute, 1997) solves this problem by augmenting the regression model with an autoregressive model for the random error, thereby accounting for the autocorrelation of the errors. The model is:

$$\tilde{Y}_t = \beta_0 + \beta_1 * (YR10_{t-1}) + \beta_2 * (MR3_{t-1}) + \beta_3 * (VOL_{t-1}) + \beta_4 * (CDATE_t) + \nu_t$$

where $\nu_t = -(\phi\nu_{t-1}) + \epsilon_t$, and $\epsilon_t \sim IN(0, \sigma_\epsilon^2)$, and \tilde{Y}_t is the de-trended Y series.

Since the four point strata behave somewhat differently, we again separate the model by these four strata groups. The modeling process is based on information obtained from combining the results in Section 3.1 and Section 3.2. For each strata, only the variable $YR10_{t-1}$ was found to be significant (marginally in strata 0), with parameter estimates as shown in Table 3.6.

Of course, these models will fit better than the pure AR(1) model of Section 3.2 (Table 3.4), since AR(1) is a subset of the autoregressive error model. Similarly, they will fit better than the WLS regression models of Section 3.1 (Table 3.3). The comparison with Table 3.3 is not completely valid for the reasons mentioned in comment (c) of the previous section: the WLS regression model uses weighted least squares to account for widely varying number of loans per month, while the pure AR(1) and autoregressive error models do not. How to resolve this problem is the main thrust of this thesis, discussed in detail in Chapter 4.

Table 3.1: Correlation Coefficients by Point Strata

Strata	YR10	MR3	TERM	VOL	CDATE
-1	0.74820 <.0001	0.16589 0.0726	0.35195 <.0001	-0.54501 <.0001	-0.35764 <.0001
0	0.84388 <.0001	0.06411 0.4867	0.55829 <.0001	-0.46209 <.0001	-0.57191 <.0001
1	0.90636 <.0001	0.18879 0.0389	0.44101 <.0001	-0.58110 <.0001	-0.53709 <.0001
2	0.90868 <.0001	0.25922 0.0050	0.46093 <.0001	-0.65969 <.0001	-0.57579 <.0001

*Top values shown in each row are correlation coefficients of possible explanatory variables with effective yield (Y). Bottom values shown in each row are two-sided p-values for testing the null hypothesis that the correlation = 0.

Table 3.2: P-Values of Full and Reduced WLS Models (All Strata Combined)

Full Model:		R-square=.7223	RMSE=.03211	SBC=-3254.61	
Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t
Intercept	1	0.04525	0.00192	23.59	<.0001
YR10	1	0.51466	0.03678	13.99	<.0001
MR3	1	0.08331	0.03313	2.51	0.0123
VOL	1	-0.02122	0.00543	-3.91	0.0001
CDATE	1	-0.00002	0.00001	-1.76	0.0783
Reduced Model:		R-square=.7181	RMSE=.03228	SBC=-3239.35	
Variable	Parameter DF	Standard Estimate	Error	t Value	PR > t
Intercept	1	0.04640	0.00174	26.62	<.0001
YR10	1	0.57057	0.02078	27.45	<.0001
VOL	1	-0.02530	0.00510	-4.96	<.0001

Table 3.3: Weighted Least Squares Regression Model Parameter Estimates

Strata	Intercept	LYR10	LMR3	LVOL	N	R-Square	RMSE*
point=-1	0.04968 <0.0001	0.48532 <0.0001		-0.02133 0.02680	14,624	0.6408	0.00285
point=0	0.0437 <0.0001	0.55027 <0.0001			27,359	0.7591	0.00227
point=1	0.02725 <0.0001	0.7809 <0.0001	0.06321 0.0121		8,7802	0.8514	0.00256
point=2	0.02111 <0.0001	0.88444 <0.0001	0.06538 0.0268		4,977	0.8805	0.00289
Combined	0.03812 <0.0001	0.64426 <0.0001			55,762	0.7474	0.00283

*Assuming $n = N/120$ in individual strata, $n = N/480$ for combined.

Table 3.4: Parameter Estimates of Pure AR(1) Model (Separate Strata)

Point Strata	Intercept	$\hat{\phi}$	R-Square	RMSE
-1	0.0781	0.8723	0.7361	0.00216
0	0.0805	0.9339	0.8670	0.00161
1	0.0855	0.9833	0.9391	0.00182
2	0.0858	0.9764	0.9271	0.00227

Figure 3.1: Residual*CDATE plot of WLS Model, Strata (-1)

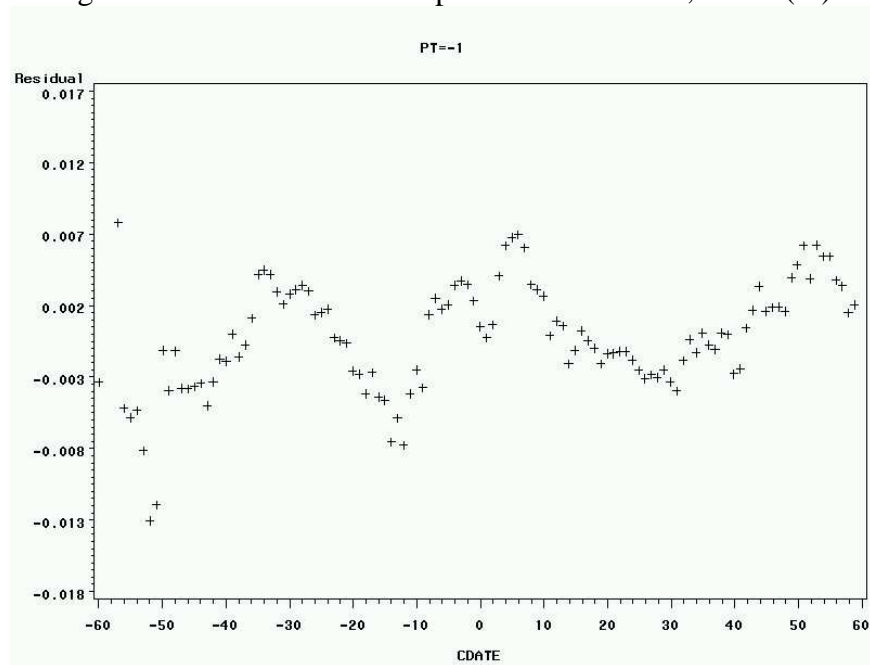


Figure 3.2: Residual*CDATE plot of WLS Model, Strata (0)

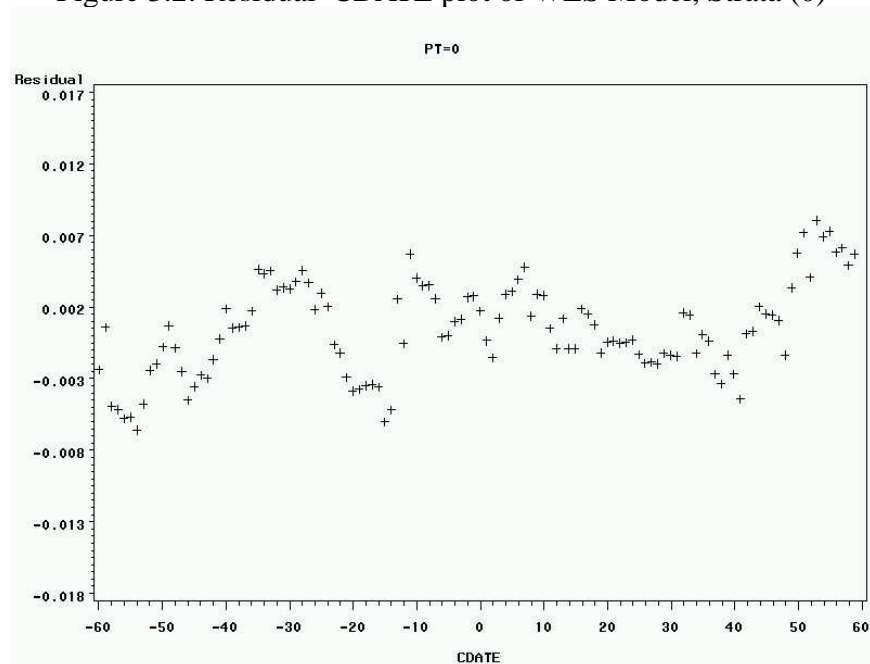


Figure 3.3: Residual*CDATE plot of WLS Model, Strata (1)

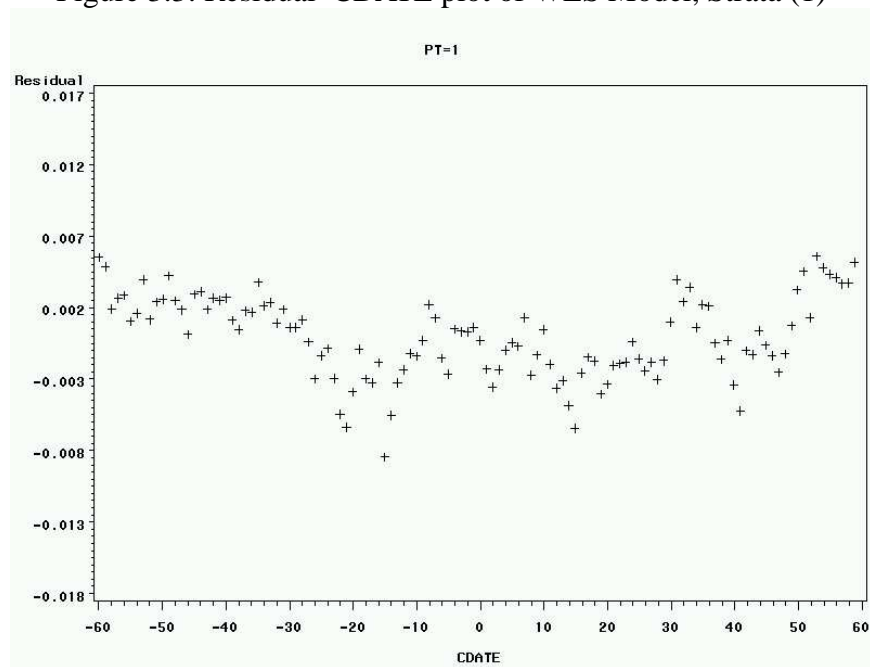


Figure 3.4: Residual*CDATE plot of WLS Model, Strata (2)

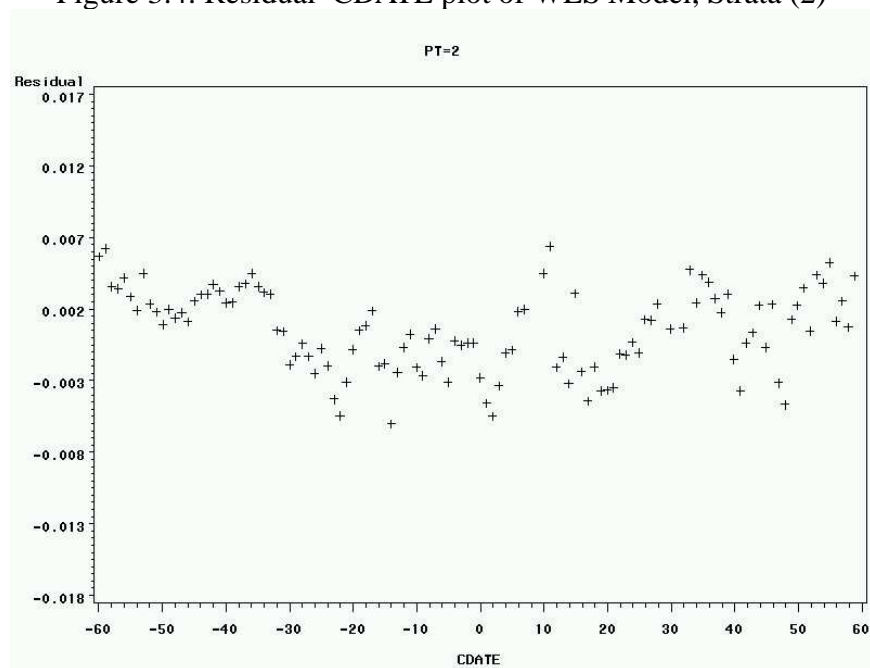


Table 3.5: Fit of AR(1) Difference Model

Point Strata	RMSE
-1	0.00218
0	0.00161
1	0.00276
2	0.00331

Table 3.6: Parameter Estimates of AR(1) Error Model

Strata	Intercept	β (YR10)	ϕ	R-Square	RMSE
-1	0.0620	0.2551	0.8544	0.7656	0.00204
0	0.0577	0.3470	0.8764	0.8978	0.00141
1	0.0781	0.1394	0.9859	0.9434	0.00176
2	0.0756	0.1774	0.9796	0.9332	0.00219

Figure 3.5: Residual*CDATE plot of AR(1) Model, Strata (-1)

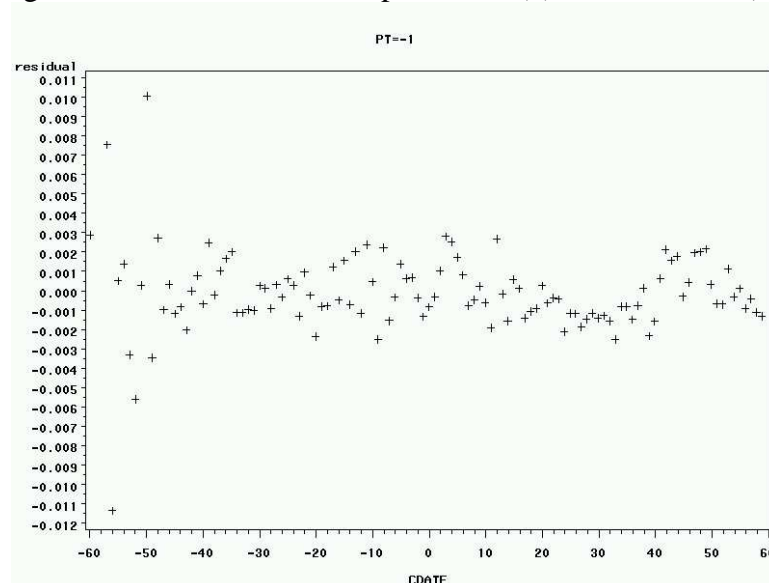


Figure 3.6: Residual*CDATE plot of AR(1) Model, Strata (0)

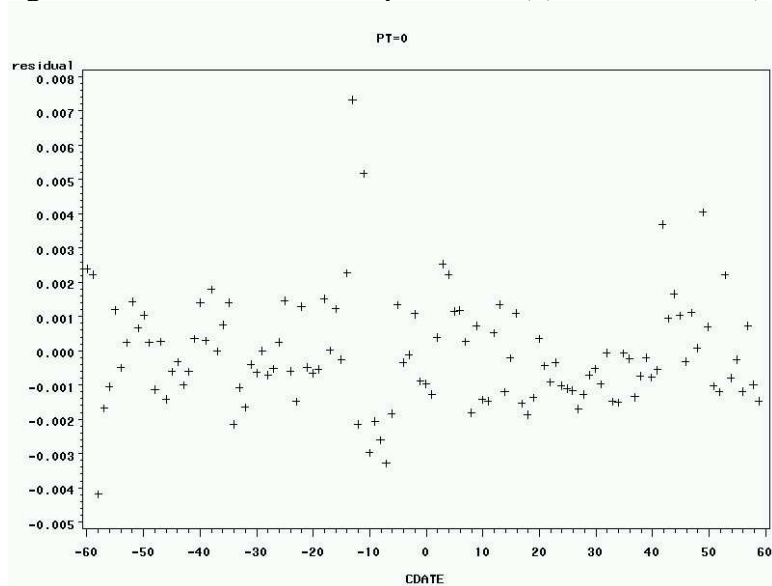


Figure 3.7: Residual*CDATE plot of AR(1) Model, Strata (1)

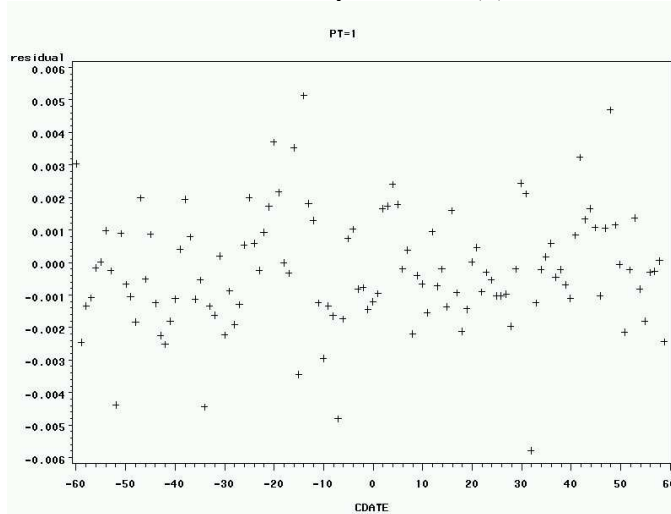
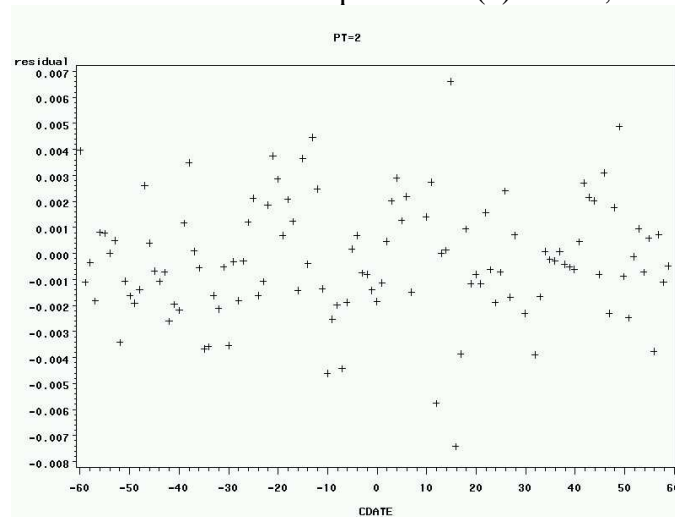


Figure 3.8: Residual*CDATE plot of AR(1) Model, Strata (2)



CHAPTER 4

WEIGHTED AUTOREGRESSIVE MODEL

As discussed in the previous chapter, the weighted regression model might be a good model to predict Y (effective yield) if the errors were independent. We previously noted that the successive records were highly correlated with one another so that an AR(1) model seems to be proper for this data set. However, the classical AR(1) model assumes that the series consists of one observation of the same unit at periodic time intervals. In fact, for the problem at hand, the same unit is not observed repeatedly. Rather, one observes some loans in one month, and other loans in other months. If one wishes to consider the unit of observation as “average value of residential loans in the state of California during a particular month”, then perhaps an AR(1) model for this average process would be appropriate. However, in this situation, we do not observe the true mean value each month. Rather, we have the mean over a sample of size n from among the loans made by a particular bank during that month. Perhaps this sample can be assumed to be a “random sample”, but it certainly is not the case that each observation has the same precision of measurement with respect to being an estimate of the true monthly mean.

First, let us review the model for a standard periodic AR(1) process with parameter ϕ . In that case (Tsay, 2002),

$$(Y_t - \mu_t) = \phi(Y_{t-1} - \mu_{t-1}) + \epsilon_t : \epsilon_t \sim IID N(0, \sigma_\epsilon^2);$$

so that

$$\epsilon_t = (Y_t - \mu_t) - \phi(Y_{t-1} - \mu_{t-1}), \text{ with}$$

$$\begin{aligned}\sigma_\epsilon^2 &= \text{VAR}(\epsilon_t) = \text{VAR}(Y_t) + \phi^2 * \text{VAR}(Y_{t-1}) - 2\phi * \text{COV}(Y_t, Y_{t-1}), \\ &= \sigma_y^2 + \phi^2 \sigma_y^2 - 2\phi(\phi \sigma_y \sigma_y) = \sigma_y^2(1 - \phi^2),\end{aligned}$$

assuming the μ_t, μ_{t-1} are known constants, $|\phi| < 1$, and $\text{VAR}(\epsilon_t)$ does not depend on t . Since each ϵ_t has the same variance, then using OLS to minimize the SSE makes sense.

So, $SSE = \sum_{t=2}^n \epsilon_t^2 = \sum_{t=2}^n [(Y_t - \mu_t) - \phi(Y_{t-1} - \mu_{t-1})]^2$, and setting $\frac{\partial SSE}{\partial \phi} = 0$ leads to:

$$\hat{\phi} = \frac{\sum_{t=2}^n (Y_t - \mu_t)(Y_{t-1} - \mu_{t-1})}{\sum_{t=2}^n (Y_{t-1} - \mu_{t-1})^2}. \quad (4.1)$$

This is the standard estimate of ϕ in the AR(1) model.

Now let us suppose that \bar{Y}_t follows an AR(1) process, where \bar{Y}_t is the mean value for month t , if one could average over all loans within a strata within month t . However, we do not actually observe \bar{Y}_t , but, rather, Y_t , which is the mean effective yield over n_t loans chosen “randomly” during month t . Thus,

$$Y_t = \bar{Y}_t + \delta_t$$

with $\delta_t \sim N(0, \frac{\sigma_a^2}{n_t})$, where σ_a is the “within-month” SD of Y values.

If the AR(1) process holds for the true mean series, then

$$\begin{aligned}(\bar{Y}_t - \mu_t) &= \phi(\bar{Y}_{t-1} - \mu_{t-1}) + \epsilon_t, \\ \Rightarrow (Y_t - \delta_t - \mu_t) &= \phi(Y_{t-1} - \delta_{t-1} - \mu_{t-1}) + \epsilon_t \\ \Rightarrow \epsilon_t &= (Y_t - \delta_t - \mu_t) - \phi(Y_{t-1} - \delta_{t-1} - \mu_{t-1}) \\ \Rightarrow e_t \equiv (Y_t - \mu_t) - \phi(Y_{t-1} - \mu_{t-1}) &= \epsilon_t + \delta_t - \phi\delta_{t-1}, \quad \text{so that} \\ \text{VAR}(e_t) &= \text{VAR}(\epsilon_t) + \text{VAR}(\delta_t) + \text{VAR}(\phi\delta_{t-1}) \\ &= \sigma_y^2(1 - \phi^2) + \sigma_a^2\left(\frac{1}{n_t} + \frac{\phi^2}{n_{t-1}}\right)\end{aligned} \quad (4.2)$$

(assuming μ_t, μ_{t-1} are known, and that δ_t, δ_{t-1} are uncorrelated with each other and with the ϵ_t 's). So, more succinctly, the first order weighted Autoregressive Model [WAR(1)]

can be written as:

$$(Y_t - \mu_t) = \phi(Y_{t-1} - \mu_{t-1}) + e_t, \quad \text{where} \quad (4.3)$$

$$e_t \sim N(0, \sigma_e^2), \quad \text{where}$$

$$\sigma_e^2 = \sigma_y^2(1 - \phi^2) + \sigma_a^2\left(\frac{1}{n_t} + \frac{\phi^2}{n_{t-1}}\right).$$

Now $\text{VAR}(e_t)$ is no longer constant, unless n is constant, $\sigma_a^2 = 0$, or $n \rightarrow \infty$. If the second term in (4.2) is insignificant compared to the first, then weighted and unweighted estimates for ϕ will be about the same. However, if the second term is substantial, then a better estimate can be found by weighting the errors inversely proportional to their variances. That is,

$$\hat{\phi} = \frac{\sum_{t=2}^n W_t (Y_t - \mu_t)(Y_{t-1} - \mu_{t-1})}{\sum_{t=2}^n W_t (Y_{t-1} - \mu_{t-1})^2}$$

where $W_t \propto \frac{1}{[(1 + \frac{\sigma_a^2}{\sigma_y^2(1-\phi^2)})(\frac{1}{n_t} + \frac{\phi^2}{n_{t-1}})]}$. Assuming one knew σ_a^2 , one could use iteratively re-weighted least squares to estimate $\hat{\phi}$ and $\hat{\sigma}_y$, using the relationship (from 4.3) that:

$$E(\sigma_e^2) = \sigma_y^2(1 - \phi^2) + \sigma_a^2(1 + \phi^2) * E\left(\frac{1}{n}\right).$$

The weighted autoregressive model [WAR(1)] displayed above will differ significantly from the classical AR(1) model if the n_t 's are quite different from one another, especially if some of the n_t 's are quite small. This is exactly the situation which we have with the loan series, since there are some months in some strata for which the number of loans (n_t) over which the average is calculated are extremely small. Treating residuals from these months as being equally as important as residuals from months with very large sample sizes (as the AR(1) model does) clearly can not be correct. Thus, the mortgage loan data series appears to be good candidate upon which to fit the WAR(1) model. The question that next arises is: "Can reliable estimates be obtained?"

The true means (μ_t^J) for each strata for each month, if there were no errors present at all, are assumed to be given by the periodic model of Section 3.2:

$$\mu_t^J = \mu^J + \alpha_\nu, \quad \text{with } t = 12K + \nu.$$

The Y_t 's are, of course, observed. We initially estimate the four μ^J by the values given in the “weighted mean” column of Table 2.1 and the twelve α_ν by the values given in Table 2.5. The μ^J estimates will be modified in the course of the iterations, since the mean weighted error must be zero under the assumptions of the model. This leaves $\hat{\phi}$, $\hat{\sigma}_a$ and $\hat{\sigma}_y$ to be estimated. If one of these three were known, the other two could be found quickly by iteration. [For pure AR(1), for example, $\sigma_a = 0$, so $W_t = 1$. This leads to a 1-step estimator of $\hat{\phi}$ using equation (4.1) of this chapter. Substituting this value of $\hat{\phi}$ into the SSE expression, one can solve for the Mean Square Error (MSE) and then solve for $\hat{\sigma}_y$ by equating $\text{MSE} = \sigma_y^2(1 - \phi^2)$, (provided, of course, that $|\phi| < 1$).] Ideally, when the time series was collected, we would have been given both the sample mean and the sample SD of the n_t loans which comprised the set evaluated within a point stratum during month t . If this were the case, we would conduct some simple analyses to determine if pooling the 120 sample variances within a stratum together to obtain a pooled strata variance made sense. Assuming that it did, we would have done so, obtaining an estimate of $\hat{\sigma}_a$ for each strata. Unfortunately, the sample SD's for each month were not recorded. In this case, there is no unique solution for $(\hat{\phi}, \hat{\sigma}_a, \hat{\sigma}_y)$ within each stratum. One can fix one of σ_a or σ_y and solve iteratively for the other two parameters. The results of doing this for each strata are shown in Table 4.1. At one extreme, σ_y was set to zero, as shown at the top of Table 4.1. At the other extreme, σ_a was set to zero, as shown at the bottom of the table. The estimates of $\hat{\phi}$ vary considerably within each stratum, depending on the assumption made: [.8067, .9620] in stratum -1, [.9198, .9769] in stratum 0, [.9398, .9545] in stratum 1 and [.9295, .9362] in stratum 2. The WMES*E6 value are the weighted mean square errors (* 10^6), so that their square roots would be estimates of the “typical” RMSE.

As $\hat{\sigma}_y$ decreases, (and $\hat{\phi}$, correspondingly, increases), one will notice that WMSE decreases. This does not mean, however, that one should choose the model at the top of Table 4.1 as the “best” model. In fact, both extremes displayed in Table 4.1 seem improbable. As mentioned in Chapter 2, it is very unlikely that the SD's for the monthly Y_t 's will

ever be obtained. The only estimate which we have, based on a few months observation in strata 0, is “ $\sigma_a = .003$ ”. This is surely a rounded value, and even if valid for strata 0, is probably not quite correct for the other strata. Nonetheless, it appears to be a reasonable compromise value and does fall into the “region of feasibility” for all four strata, as can be noted from Table 4.1. If one fixes $\sigma_a = .003$ in all four strata, the WAR(1) parameter estimates are as shown in Table 4.2 .

The RMSE’s displayed in Table 4.2 for the WAR(1) model are substantially smaller, for each strata, than those given for the AR(1) model of Table 3.4. However, just as the pure AR(1) model of Table 3.4 could be improved by including the regressor YR10, as shown in Table 3.6, we should check the weighted residuals given from the WAR(1) models to see if they can be further explained by including regressor terms. That is, we consider the Weighted Autoregressive(1) Error [WAR(1)E] Model:

$$(Y_t - \mu_t) = \phi(Y_{t-1} - \mu_{t-1}) + [\beta_0 + \beta_1 * YR10_{t-1}] + e_t$$

with $VAR(e_t)$ given by equation (4.2). For strata -1 and 0, including the YR10 factor was only marginally significant, but for strata 1 and 2, the β_1 coefficient was more significant. The final parameter estimates for β_0 , β_1 and RMSE for each strata under the WAR(1)E model are shown in Table 4.3. In each of the models, the $\hat{\mu}^J$, $\hat{\sigma}_y$, $\hat{\sigma}_a$ and ϕ of Table 4.2 are re-used. Technically, one should solve iteratively for these, along with β_0 and β_1 , but this was not done here.

As expected, the RMSE’s from this final model are only slightly smaller than those from the WAR(1) model of Table 4.2. Thus, while $YR10_{t-1}$ does lend some slightly significant information to the prediction of effective yield at time t , the great majority of the variability in effective yield can be explained by regressing the variable on itself, appropriately weighted, as the WAR(1) model does.

Table 4.1: Estimates of σ_y , σ_a and ϕ for WAR(1) Model

Strata	J=-1	J=0	J=1	J=2
avg(1/n)	.06492	.01437	.03700	.08585
avg(n)	121.87	227.99	73.35	41.48
μ^J	.07464	.07622	.07850	.07989
μ^J	.06783	.06732	.07399	.07968
σ_y	.00000	.00000	.00000	.00000
σ_a	.00286	.00611	.00513	.00428
ϕ	.9620	.9769	.9545	.9295
WMSE*E6	1.0257	1.0490	1.8605	2.9335
μ^J	.07148	.07016	.07444	.07933
σ_y	.00100	.00100	.00100	.00100
σ_a	.00311	.00629	.00513	.00425
ϕ	.9539	.9709	.9529	.9320
WMSE*E6	1.2852	1.1606	1.9481	3.0332
μ^J	.07478	.07388	.07534	.07911
σ_y	.00200	.00200	.00200	.00200
σ_a	.00310	.00617	.00496	.00410
ϕ	.9350	.9577	.9493	.9345
WMSE*E6	1.6723	1.2803	2.1228	3.2129
μ^J	.07646	.07685	.07625	.07912
σ_y	.00300	.00300	.00300	.00300
σ_a	.00238	.00483	.00441	.00377
ϕ	.9038	.9356	.9456	.9357
WMSE*E6	2.3180	1.7498	2.3157	3.4086
μ^J	.07683	.07895	.07719	.07921
σ_y	.00325*	.00400	.00400	.00400
σ_a	.00184	.00084	.00328	.00320
ϕ	.8799	.9200	.9423	.9362
WMSE*E6	2.7747	2.4763	2.5434	3.6312
μ^J	.07698	.07899	.07798	.07989
σ_y	.00336	.00403	.00506	.00589
σ_a	.00000	.00000	.00000	.00000
ϕ	.8067	.9198	.9398	.9311
WMSE*E6	3.9493	2.5058	2.9925	4.6219

Table 4.2: Parameter Estimates for WAR(1) Model, $\sigma_a = .003$

Strata	J=-1	J=0	J=1	J=2
μ^J	.07531	.07817	.07736	.07924
σ_y	.00227	.00358	.00417	.00427
σ_a	.00300	.00300	.00300	.00300
ϕ	.9293	.9232	.9418	.9362
WMSE*E6	1.7927	2.1247	2.5912	3.7016
RMSE	.00134	.00146	.00161	.00192

Table 4.3: Estimates of β_0 , β_1 and RMSE for WAR(1)E Model

Strata	J=-1	J=0	J=1	J=2
β_0	-.00179	-.00239	-.00201	-.00230
β_1	.02887	.03758	.03158	.03586
RMSE	.00133	.00143	.00159	.00191

CHAPTER 5

CONCLUSION

In this thesis, we have examined the predictability of the effective yield of residential mortgage loan rates as a function of economic variables which would be available to a loan officer preparing to set a loan rate. Here are some points which have been revealed from the analyses:

1. The mean effective yield does increase linearly with the point class, but the rate of increase per increase in points is about .0018 rather than .01. That is, the effective yield (contract rate plus points) is fairly similar for those in different point classes, but, on average, every increase in point class is penalized by an .0018 increase in effective yield rate.
2. There do not appear (over the 10-year period studied) to be sustained linear trends in the loan rate. However, within each of the four point strata, there did appear to be consistent small but statistically significant periodic variations such that the loan rate was really lowest in December and January and highest in June and July. One might think that this occurs because there are more houses sold in the summer, when families with children are more likely to move, so that market conditions are good for sellers. However, the data do not reveal any significant increase in the number of loans made during summer months; just a periodic increase in the rate charged. Also, the $PT = 0$ strata, which is the largest, shows a secondary maximum in September which is not observed in the other strata (see Figure 2.4).

3. Of the possible explanatory regressor variables, the 10-year interest rate ($YR10_t$) is most highly correlated with effective yield (Y_t), with correlation coefficient varying from $r=.75$ for strata -1 to $r=.91$ for strata 2. However, $YR10_t$ would probably not be available to a loan officer at time t , so s/he would need to employ $YR10_{t-1}$. This is also very highly correlated with Y_t , but not as strongly as $YR10_t$ is. In addition, after one considers regressing Y_t on itself (autoregression), the additional explanatory ability of $YR10_{t-1}$, while statistically significant in some strata, is very small.

4. Ignoring either the unbalanced monthly sample size problem or the problem of correlated errors would lead to substantially larger RMSE's than would be found with a correct model. Collecting the RMSE values from the WLS model (Table 3.3), the AR(1) model (Table 3.4), the AR(1)E model (Table 3.6), the WAR(1) model (Table 4.2) and the WAR(1)E model (Table 4.3) yields Table 5.1. From this, we see that accounting for correlated errors (as any of the latter four models do) yields a very substantial improvement over the WLS models. Once autocorrelation has been accounted for, including $YR10_{t-1}$ as a regressor causes approximately an 8% decline in RMSE for the AR(1) models, but only a 1% decline in RMSE for the WAR(1) models. Some of the apparent explanation in the AR(1) \rightarrow AR(1)E class is probably spurious, with the regressor variable compensating for some error caused by treating the time series as if every observation was of equal precision.

5. The major new statistical idea introduced in this thesis is that of the first order Weighted Autoregressive Model [WAR(1)], formally defined in equation (4.3) of Chapter 4. This model states that the true mean monthly series can be modeled as an AR(1) process, but that the observed series can not, since it is constructed by taking an average of n_t values at each time t . For the classical time series, one could either view n_t as infinite or say that $n_t = 1$, but $\sigma_a = 0$. Although the two explanations appear quite different, they express the same idea - in the classical AR(1) model,

Table 5.1: Comparison of RMSE for Five Different Models

Point Strata	WLS	AR(1)	AR(1)E	WAR(1)	WAR(1)E
-1	0.00285	0.00216	0.00204	0.00134	0.00133
0	0.00227	0.00161	0.00141	0.00146	0.00143
1	0.00256	0.00182	0.00176	0.00161	0.00159
2	0.00289	0.00227	0.00219	0.00192	0.00191

there is no other source of variability; the observations follow an AR(1) model. If the sample sizes (n_t) are equal or nearly equal, the WAR(1) model has no real advantage over the standard AR(1) model. In this case, assuming σ_a were known, it would be possible to decompose the error into that part due to the AR(1) process and that part due to the sampling error, but the estimates of $\hat{\phi}$ would be the same, since the weights would be approximately constant. However, in a case such as was examined in this thesis, where the n_t 's vary widely, the effects can be appreciable, as can be observed from Table 5.1. Such cases are not infrequent in practice.

6. The most serious limitation of the data used in this thesis is that the raw data from which the monthly series were constructed were not available to us. Hence, we were not able to estimate σ_a precisely within each strata. If we had been able to do so, it is likely that the estimates of σ_a would have varied somewhat by strata, perhaps slightly increasing or decreasing the RMSE's shown for the WAR(1) models in the last two columns of Table 5.1.
7. Although the modeling effort appears to have been fairly successful, there are several caveats. One is that we were assuming that the n_t loans sampled in month t were a random sample of all loans in this strata during month t . Considering that these data were collected by a bank from its internal records, it would not be at all

surprising if the loans were really generated by a sampling process more similar to a cluster sample, where all the loans made by certain loan officers in a certain month were included in the sample. This would tend to deflate estimates of $\hat{\sigma}_a$ and might inflate $\hat{\sigma}_y$. A more serious concern is that the strata are viewed as being independent of one another, but that is probably not the case. Depending on the economic conditions, the proportion of loans in each of the four strata varied considerably. Thus, the mean strata effects (μ^J) were probably not really constant over the period examined.

8. Finally, and perhaps of most interest to statisticians, the procedure used in this thesis to obtain the WAR(1) estimates is based on iteratively reweighted least squares methods. A more elegant, general procedure would be to use maximum likelihood methods. One suspects that the results would be similar, although as $|\phi|$ approaches 1, as was the case here, it is well-known that the methods could yield dissimilar results. We will leave examination of this question as a topic for future research.

BIBLIOGRAPHY

- [1] Clauretie, Terrence M. (1999). *Real Estate Finance: Theory and Practice*, Third Edition, Prentice Hall, Upper Saddle River, NJ.
- [2] SAS Institute (1995). *SAS/STAT User's Guide*, Version 6, Fourth Edition, Cary, NC.
- [3] SAS Institute (1997). *SAS/ETS User's Guide*, Version 6, Second Edition, Cary, NC.
- [4] Shumway, Robert H. (2000). *Time Series Analysis and its Applications*, Springer, New York.
- [5] Tsay, Ruey S. (2002). *Analysis of Financial Time Series*, Wiley, New York.
- [6] Wahl, Kenneth L., Thomas, Wilbert O. and Hirsch, Robert M. (1995). *Stream-Gaging Program of the U.S. Geological Survey*, US Geological Survey Circular 1123, Reston, Virginia,.